



Forskeres selektionsstrategier I informationssøgning under "overload" betingelser

Hjørland, Birger

Published in:
Skrifter om Anvendt og Matematisk Lingvistik

Publication date:
1985

Document version
Også kaldet Forlagets PDF

Citation for published version (APA):
Hjørland, B. (1985). Forskeres selektionsstrategier I informationssøgning under "overload" betingelser. *Skrifter om Anvendt og Matematisk Lingvistik*, 11, 69-104.

FORSKERES SELEKTIONSSTRATEGIER I INFORMATIONSSØGNING UNDER
"OVERLOAD"-BETINGELSER

Birger Hjørland

INDLEDNING.

I videnskabelig forskning og tilsvarende situationer står man ofte overfor det problem, at litteraturmængderne er så store, at de er uoverkommelige for forskeren indenfor den til rådighed stående tid og med de herskende normer, idet man jo ikke blot skal læse andres forskning, men også selv producere ny viden. Man taler på engelsk om "information input overload", jeg vil her benytte termen informationsoverlæs(ning) eller blot overlæsning.

I et tidligere arbejde (Hjørland, 1984a) har jeg systematiseret de faktorer, der efter min opfattelse påvirker informationsoverlæsningen i fire kategorier:

I. Faktorer knyttet til forskernes informationsbearbejdning, herunder især deres litteratursøgningsstrategier.

II. Faktorer knyttet til forskernes opgaveformulering og -løsning. Herunder det indlysende forhold, at forskere ofte vælger opgaveemner, hvorom der er grund til at formode, at litteraturarbejdet er overkommeligt, hvilket kan have den negative konsekvens, at forskningen koncentrerer sig om perifere emner, fordi disse er lettest at gå i lag med. Under dette punkt hører også hele spørgsmålet om den videnskabelige specialisering, hvor forholdet imidlertid ofte er, at en meningsfuld specialisering først kan foretages når man har fordøjet en omfattende litteratur, fordi emnet udgør en helhed, der ikke lader sig opdele vilkårligt. Hvis man alligevel gør dette, bliver resultatet ofte præmaturt eller pseudospecialisering, hvilket igen øger overlæsningen for andre forskere.

III. Faktorer, der er knyttet til fagområdet, f.eks. omfanget af forskningslitteraturen på område, hvor klart den er struktureret etc.

IV. Faktorer, der er knyttet til forskerens sociale situation, f.eks. hvorvidt han selv kan strukturere sin forskningstid, hans adgang til medhjælp og ressourcer, andres interesse i at holde ham informeret o.s.v.

I denne artikel vil vi tage hul på en realbehandling af nogle af de konkrete forhold, der påvirker overlæsningen. Vi vil indskrænke os til forhold omkring forskeres informationssøgning og lade andre forhold ligge til evt. senere arbejder. Et overblik over de behandlede faktorer kan fås gennem indholdsfortegnelsen (s. F33).

FAKTORERNES SYSTEMATISERING.

Informationssøgning er en arbejdsproces, og graden af overlæsning hænger naturligvis i høj grad sammen med, hvordan denne arbejdsproces tilrettelægges, herunder hvordan litteratursøgningen kombineres med den øvrige forskningsproces (se Clod Poulsen, 1975). Såfremt man især indskrænker litteratursøgningen til en kortvarig fase og på forhånd bestemmer, hvilken litteratur der er relevant, så oplever man naturligvis umiddelbart en meget mindre overlæsning end hvis man vedbliver med at søge, under søgningen benytter enhver ny viden om søgemuligheder til at udvide søgningen og enhver ny erkendelse om emnet til at ændre sine søgekriterier. I tidsbegrænsede projekter og ved studenterarbejder kan "one step" søgninger være nødvendige, og brugeren må derfor på forhånd fastsætte kriterier for, hvad der anses for relevant (a priori definition på relevans). I mere seriøs og dybtgående forskning kan man imidlertid ikke klare sig med en sådan nødløsning, og meget af den psykiske belastning i forskning ligger vel netop i det forhold, at det, man allerede har investeret megen tid og kræfter i at opspore, pludselig viser sig mindre frugtbart end et helt nyt spor, hvor man i søgeprocessen (a posteori) ændrer sine selektionskriterier. Den midler-

tidige lettelse i informationsoverlæsningen man opnår ved a priori at definere sine relevanskriterier, vil vende sig til en øget belastning senere. Dette under forudsætning af, at man fastholder en videnskabelig problemstilling og ikke f.eks. skifter målsætning og udarbejder en bibliografi eller "review" over området.

Netop fordi vi forudsætter en fastholdelse af egentlige videnskabelige målsætninger eller erkendelsesfremskridt, bliver de afgørende forhold omkring informationssøgningen muligheden for at kunne fastslå kvaliteten af et videnskabeligt arbejde uden at skulle foretage en fuldstændig gennemlæsning og afprøvning af ethvert værk. Det er naturligvis forskelligt fra sag til sag hvilke "cues", der i særlig grad røber et arbejdes kvalitet. I nogle tilfælde kan man komme på sporet af et væsentligt arbejde gennem en anmeldelse, i andre tilfælde via kendskab til forfatterens tidligere arbejder o.s.v. Generelt må man imidlertid kunne sige, at selve teksten er den bedste indikator for et arbejdes kvalitet, kendskab til forfatteren den næstbedste indikator og kendskab til emnet eller disciplinen den trediebedste. Tilsvarende kan man sige, at direkte viden om disse forhold er bedre end andenhåndsviden, d.v.s. viden om hvad f.eks. anmeldere eller andre forfattere mener. Udfra disse overvejelser vil jeg mene, man kan opstille seks punkter fra relativt gode og sikre kriterier til relativt usikre og tilfældige kriterier:

1. Selektion efter kriterier, der bygger på direkte kendskab til et arbejdes kvalitet, f.eks. checks af, hvordan forfatteren behandler emner, som brugeren på forhånd kender noget til, hvorvidt han medtager de teorier, som brugeren anser for væsentlige o.s.v.
2. Et kendskab til forfatteren, hans videnskabelige og faglige baggrund, produktion m.v.
3. Indirekte kendskab til værket eller forfatteren, herunder f.eks. hvem der citerer værket eller forfatteren, hvem der er udgiver (kendt forlag, offentlig institution eller andet), såfremt det er en artikel, er den da publiceret i et anerkendt tidsskrift? Hvordan er anmeldelserne? Hvem anbefaler det? Kommer det i mange udgaver/oplag og er det stadig i handelen?

4. Kendskab til forskningsområdet og dets status.
5. Afgrænsning i tid.
6. Afgrænsning på sprog, form, udstyr, sted, tilgængelighed.

Denne disposition giver anledning til utallige spørgsmål, som stort set er ubesvarede, idet biblioteks- og informationsvidenskaben - som også Patric Wilson (1983) konstaterer - ikke har beskæftiget sig synderligt med kvalitetsproblemer, idet man stort set er tilbøjelig til at opfatte biblioteket som en telefoncentral, hvor teknikernes opgave er den effektive kommunikation uden tanker på indholdet. Jeg mener - med Wilson - at det er vigtigt for informationsvidenskaben at få taget sådanne kvalitative problemer op. Udforskningen af, hvorvidt ovenstående kriterier rent faktisk er gode indikatorer for kvalitet er nærmest af videnssociologisk art (og meget vanskelige, idet de forudsætter en evaluering af enkelte forskningsområders tilstand). Udforskningen af, hvordan forskere bruger disse kriterier, herunder deres holdninger og viden om disse spørgsmål er snarere et psykologisk spørgsmål. Det er imidlertid min opfattelse, at man ikke opnår ret meget ved en psykologisk udforskning af forskernes adfærd, viden og holdning, hvis man ikke samtidig interesserer sig for de virkelige forhold. Det er vigtigt og interessant at udforske, hvordan de enkelte muligheder (f.eks. boganmeldelser) faktisk udnyttes af forskere, og hvad forskernes viden og holdning er til f.eks. de største anmeldelsestidsskrifter. En sådan forskning bør imidlertid ikke stå alene, men suppleres med en undersøgelse af anmeldelsernes produktions- og distributionsforhold og af deres faktiske værdi.

DIREKTE CHECK AF PRIMÆRDOKUMENTER.

Direkte bladren i bøger og gennemsyn af tidsskrifter er et vigtigt kriterium for fastsættelse af en kildes værdi, og vi må derfor antage, at forskere, der har god adgang til dette, f.eks. ved cirkulationsordninger på deres institut, er mindre udsatte for overlæsning end forskere, der i højere grad må benytte sekundære hjælpemidler til litteraturudvælgelse.

Eet af de mest benyttede selektionsprincipper er, at man undersøger, hvorvidt forhold man selv kender særlig godt, er behandlet på en tilfredsstillende måde. Professor Tranekjær Rasmussen nævnte ved en forelæsning engang, at han altid undersøgte, hvordan amerikanske psykologibøger omtalte Edgar Rubin. Bibliotekarere vil typisk teste udenlandske leksika på, hvordan de omtaler Danmark. Iøvrigt undersøger man, om hovedteorierne på området er repræsenteret, om der forekommer store tilfældige udeladelser i behandlingen, om referencelisten virker dækkende i forhold til de forfattere, man selv værdsætter etc.

Der kan naturligvis nævnes mange andre kriterier for selektion af primærdokumenter, herunder om emnet sættes i en historisk kontekst, metode o.s.v. Der findes endnu kun sparsom litteratur om det at læse/checke videnskabelig litteratur, f.eks. Riegelman (1981).

I de fleste tilfælde er der imidlertid så store praktiske vanskeligheder i at nå frem til at have primærdokumenterne i hånden, at de helt dominerende udvælgelseskriterier (kvantitativt set) ligger på indirekte metoder. De fleste forskningsområder er så omfattende, at forskeren ikke ville kunne foretage en direkte inspektion af alle dokumenter, selvom han anvendte al sin tid på dette. Der må derfor først ske en grovsortering efter indirekte metoder.

DIREKTE KENDSKAB TIL FORFATTEREN.

Næst efter værket eller artiklen selv er naturligvis kendskab til dets forfatter af betydning for at kunne konstatere et arbejdes værdi. En ellers god forfatter kan naturligvis begå en dårlig bog, så der er ingen garanti, men ikke desto mindre vil de fleste med god ret være tilbøjelige til at stifte bekendtskab med en ny bog af en forfatter, de tidligere har haft stor glæde af. En forfatter må formodes stadig at have kendskab til de forhold, han har belyst tidligere, og anvende denne indsigt på nye områder.

Direkte kendskab til en forfatter spiller en større rolle i nogle fag end i andre. Især i humanistiske fag er det almindeligt at have yndlingsforfattere. Indenfor et fag som psykologi er det i nogle retninger f.eks. psykoanalytiske, almindeligt at forbinde sin interesse med bestemte forfattere (f.eks. Freud, Jung, Adler, Fromm etc.), mens det samme er sjældent indenfor eksperimental-psykologi, uden at forfatteren dog her af den grund er betydningsløs som selektionsfaktor i litteratursøgning. Naturligvis kan man mene, at der er noget "prævidenskabeligt" over en sådan personcentreret forskningsmetode, fremfor en mere dominerende emnemæssig eller indholdsmæssig afgrænsning. Der kan imidlertid ofte i de humanistiske fag være tale om, at et forfatterskab udgør en organisk helhed, som det er vigtigt at forstå. I disse tider kan der være en tendens til at tage enkeltbegreber ud fra overordnede teorier og studere disse løsrevent. Indenfor psykoanalyse udgør f.eks. narcissisme et sådant modebegreb. Hvis man ikke har forstået klassikernes grundlæggende begreber, menneskeopfattelse og filosofi, er der en fare for at forskningen snarere end at være prævidenskabelig bliver uvidenskabelig og forårsager overlæsning hos andre forskere. (Der kan så være tale om, at andre "genopdager" klassikerne og danner en ny skole på denne baggrund, f.eks. Lacan-skolen i Paris).

Hvis man ikke på forhånd kender en forfatter, kan man lægge vægt på hans data. En fagleder i naturvidenskab fortalte mig engang, at han fortrinsvis læste forfattere, der havde arbejdet med samme emne i flere år og jævnligt publiceret herom. Så kunne man håbe, at forfatteren var begyndt at finde ud af, hvad det hele gik ud på. Denne strategi må ses på baggrund af nogle karakteristiske forhold ved megen moderne videnskab, herunder at en meget stor del udføres af meriteringsgrunde uden sigte på forskningskarriere, men som springbræt til f.eks. læge- eller psyko-logstillinger. Garvey og Griffith (1971, 352-354) dokumenterer således, hvorledes "produktive forskere" (defineret som en forsker, der producerer mindst een artikel årligt) udgør mindre end 10%, og at 65% af forfatterne i en femårsperiode kun blev citeret een gang i Annual Review of Psychology. En alvorlig svaghed ved

denne metode er imidlertid, at nye "paradigmer" i forskningen som Kuhn har påpeget især slår igennem hos unge forskere. Man risikerer således en stor grad af konservatisme ved at benytte forskeres alder/produktion som selektionskriterium.

Alt i alt må vi nok fastslå, at selektion på baggrund af kendskab til forfattere spiller en kvantitativt underordnet rolle, idet man normalt ikke i større omfang kan basere sin forskning på forfattere, man kender i forvejen.

INDIREKTE KENDSKAB TIL VÆRK ELLER FORFATTER.

Når forskerens kapacitet til selv at evaluere et værk eller en forfatter er begrænset må han støtte sin selektion på andre menneskers vurdering. Disse andre mennesker vil især være forskere selv, men kan også være f.eks. udgivere eller anmeldere udenfor forskernes kreds. I almindelighed er det ikke tilstrækkeligt at det er en forsker, det skal også være en anerkendt forsker eller en forsker, hvis synspunkter man selv respekterer/deler. Da det imidlertid kan være svært at få autoritative evalueringer, må man i mange praktiske situationer basere sin selektion på noget mindre optimale andenhåndsvurderinger. I praksis spiller især to metoder her en stor rolle: 1) at basere sin selektion på meget citerede værker/forfattere 2) at basere sin selektion på meget udbredte værker, f.eks. "bestsellers".

Mest citerede forfattere.

En af de mest udbredte metoder som selektion i litteratursøgning er sandsynligvis, at forskeren i forbindelse med såkaldt "sneboldsøgning" lægger mærke til hvilken litteratur, der går igen i referencelisterne og altså citeres mange gange. Det forhold, at mange forskere - tilsyneladende uafhængigt af hinanden - citerer samme dokument, tages som et indicium på, at dette dokument af andre er fundet relevant og evalueret positivt - og derfor værd at beskæftige sig med. D.v.s. der anlægges den betragtning, at videnskabsmændenes efterspørgsel (målt ved antal henvisninger) udgør en markedsmekanisme, og denne markedsme-

nisme antages mere eller mindre effektivt at kunne skelne god fra dårlig kvalitet og undertrykke støj.

Hvorvidt denne metode rent faktisk er effektiv afhænger af, hvordan disse markedsmekanismer fungerer. Her kan peges på bestemte forhold, der kan bevirke, at visse hensyn tilgodeses fremfor andre, og at det ikke alene er kvalitetsmæssige forhold, der afgør, hvad der bliver citeret. Som tidligere nævnt er en meget stor del af den forskning, der i dag udføres meriteringsforskning, f.eks. i forbindelse med disputatser. Det er dermed forskning, der skal leve op til akademiske standarder og være gennemførlig indenfor et kortere tidsmål. Indenfor nogle forskningsområder er en sådan meriteringsforskning helt i overmål. Det betyder ikke blot, at det er relativt unge og uprøvede forskere, der tegner "markedet" fremfor seniorforskerne, det betyder også, at bestemte typer af dokumenter har en langt større chance for at blive citeret end andre, nemlig sådanne dokumenter, der indeholder en idé, forskningsmetode eller lignende, der let lader sig anvende på et nyt emne. Det er således ikke dokumenter med den største erkendelsesværdi, men snarere med den største pragmatiske værdi, som et sådant marked vil fremavle.

Man kan tænke på mange andre former for "usaglige" hensyn, der bevirker en skævhed i hvilke forfattere, der især citeres. Man kan sågar tænke sig, at særligt provokerende eller dårlige forfattere citeres meget for at blive imødegået (hvilket har været diskuteret i forbindelse med Arthur R. Jensens artikel om race og intelligens). Metoden har også en kraftig selvforstærkende effekt: De forfattere, som i forvejen bliver stærkt citeret vil yderligere blive stærkt citeret alene fordi de er så kendte, at man vanskeligt kan gå uden om dem.

Metoden fremmer indavl, populisme og modebevægelser. Alligevel må den nok siges at være en ret uundværlig metode til at undgå overlæsning, og nogle af de nævnte ulemper kan neddæmpes ved bevidsthed om hvem der citerer, og analyser af deres motiver til at citere.

(Iøvrigt indebærer anvendelsen af denne metode en kraftig sekundærgevinst. Hvis et givet dokument trods mange henvisninger skulde vise sig at være af tvivlsom værdi, så berettiger alene det forhold, at det drejer sig om et indflydelsesrigt dokument det til en kritik, hvorimod der ikke er tilsvarende bonus at hente i at kritisere et dokument, som ingen indflydelse har.)

Mest publicerede/efterspurgte bøger.

Foruden det marked, der består af forskernes præferencer, således som det viser sig i den citerede litteratur, er der et andet marked, forlagsmarkedet, der bestemmer, hvad der skal udgives. En amerikansk informationsspecialist (Mr. Pratt) fortalte mig engang, at hans foretrukne litteratursøgningsstrategi var at søge bøger via emneregisteret til den amerikanske boghandlerkatalog "Books in Print". Det var hans opfattelse, at bøger, der var noget værd via markedets efterspørgsel ville blive udbudt til salg. Dette ville også gælde ældre værker eller mere ukurante ideer, der enten ville komme i genoptryk eller hvis væsentlige ideer ville blive indarbejdet i andre værker.

Der findes varianter af denne form. Det er ikke ualmindeligt, at forskere støtter sig på bøgernes oplag/udgave antal. Mængder af bøger kommer kun i eet oplag og forsvinder uden at sætte varigt spor. De værker, der formår at komme igen i nye udgaver har overlevet en vis prøvelse.

Om forlagsmarkedet må man nok sige, at det ikke har mulighed for at fungere så effektivt som forfatterernes citationsmarked som middel til kvalitetskontrol. Dels er forlagernes udbud langt mere selektivt end det videnskabelige marked, med målgrupper der er delvist forskellige. Dels kan forlagsfolk ikke have samme fagindsigt som videnskabsmænd. Selvom udbuddet af bøger er kolossalt, skal man ikke lade sig forlede til at tro, at specielle ting af værdi er garanteret en plads. Som forskningsbibliotekar kan man gøre den iagttagelse, at en stor del af udlånet af faglitteratur er af bøger, der ikke længere kan fås i boghandelen. Metoden vil - hvis man tror på den - naturligvis betyde en praktisk lettelse

og en begrænsning af informationsoverlæsningen, men kan nok især siges at være rimelig, hvis man ikke arbejder på et for snævert speciale eller hvis man ikke har behov for at følge forskningen i første række.

I forbindelse med forlagsmarkedet kan man skelne mellem private forlag, offentlige forlag, faglige sammenslutninger og lærde selskaber m.v. Der er fra nogle sider udtrykt stor angstelse for, hvad en øget kommercialisering af det videnskabelige forlagsmarked vil komme til at betyde for den videnskabelige kvalitet. Jeg skal her nøjes med at efterlyse nærmere undersøgelser og analyser af dette forhold.

Andre indirekte indikatorer for værks/forfatters kvalitet.

For artiklers vedkommende er det en uhyre udbredt praksis, at man evaluerer deres værdi ved at se på, om det er et anerkendt tidsskrift, der har publiceret dem. Lægevidenskabelige nyheder har således en særlig autoritet, hvis de f.eks. er publiceret i "The Lancet". Da moderne edb-baseret litteratursøgning omfatter et meget bredt spektrum af tidsskrifter, er der altså tale om, at denne selektionsmekanisme til dels sættes ud af spillet, selvom det principielt er muligt at afgrænse et emne og forlange at det skal være behandlet i bestemte tidsskrifter, hvis titler eller numre kan være lagret i en permanent søgeprofil. Der er ingen tvivl om, at mange forskere opfatter et tydeligt hierarki i tidsskrifternes kvalitet, og at dette kriterium spiller en stor rolle i begrænsningen af informationsoverlæsning. Hele spørgsmålet om videnskabelige tidsskrifters redaktion og "referee"-systemet er et vigtigt emne at udforske, og der er da også taget hul herpå (se f.eks. Wilson, 1983, p.53).

Utallige brugerundersøgelser har vist, at uformel kommunikation spiller en afgørende rolle i videnskaben. Denne uformelle kommunikation indeholder naturligvis også udveksling af synspunkter på værker og forfattere. De brugergrupper, der har mulighed for at indgå i gode uformelle netværk eller "invisible colleges"

har således anderledes gode muligheder for at undgå informationsoverlæsning end de, der alene må støtte sig på de trykte medier.

Endelig kan omtales boganmeldelser og "literature reviews" som vigtige hjælpemidler til at få udskilt kvalitetslitteraturen. Der er ikke megen tvivl om, at disse medier er særdeles populære for brugeren, men ikke særligt meriterende eller profitable for forfattere. Derfor er der en mangel på kvalitetsanmeldere og -"reviewere". Dårlige anmeldelser og reviews kan på sin side øge overlæsningen og give en falsk tryghed, der især giver problemer på længere sigt. Der er brug for bedre understøttelse af denne slags arbejde, samt for opstilling af faglige normer og retningslinier for gode litteraturoversigter.

KENDSKAB TIL FORSKNINGSOMRÅDET.

Større eller mindre kendskab til et forskningsområde, ligesom mere eller mindre underbyggede holdninger eller fordomme, kan influere på, hvad forskeren i informationssøgningen anser for relevant eller værdifuldt. Det er eksempelvis meget almindeligt at tillægge naturvidenskaberne en så stor autoritet, så dokumenter, der i ydre form, metode m.v. ligner naturvidenskabelige dokumenter på forhånd tillægges større værdi, selvom de måske vedrører et emne, hvor en naturvidenskabelig metode ikke - eller endnu ikke - er optimal.

Mest udbredte emner og synspunkter.

Videnskaberne er - hvad enten man kan lide det eller ej - til en vis grad præget af modeluner. Jo flere forskere, der slutter op bag en bestemt mode (metode, paradigme o.s.v.), desto sværere bliver det at stå udenfor. Det bliver sværere at hævde, at alle de andre tager fejl, og man selv har ret. Der kan derfor i vurderingen af videnskabelig litteratur være en form for tro på markedskræfterne, der adskiller sig fra de tidligere omtalte former for markedskræfter (mest citerede bøger, mest publicerede bøger) ved direkte at gå på emner, metoder og paradigmer.

Indenfor amerikansk psykologi har faktoranalysen i årtier været en meget dominerende statistisk analysemetode trods massiv kritik fra statistisk side, f.eks. den danske statistiker Georg Rasch. På tilsvarende måde kan der være modebevægelser i europæiske skoler, f.eks. Frankfurterskolen, Hannoverskolen o.s.v.

D.v.s. at forskere - som et middel til at afgrænse deres litteraturstudier - kan være tilbøjelige til at slutte op omkring de emnekategorier, der i forvejen er mest dominerende i bl.a. de trykte informationskanaler, de benytter sig af. Samtidig vil den samme holdning være tilbøjelig til at mistænkliggøre små mindretalsafgørelser, som trods årelang påvirkning ikke har ladet sig kue. Videnskabshistorien viser naturligvis med al tydelighed, at sandheden ikke kan afgøres ved flertalsafstemninger, hvorfor metoden er uantagelig for den, der overhovedet vil forsøge at danne sig en selvstændig opfattelse. Alligevel må de fleste nok erkende, at de i større eller mindre grad bukker under for den - og det er ikke i samme grad berettiget at kræve en stor grad af selvstændighed hos en studerende som hos en seniorforsker. Hos den første må den nok i et vist omfang anses for en legitim måde at begrænse sin overlæsning på.

Litteratur fra bestemte discipliner.

Mange emner kan behandles "tværfagligt", d.v.s. i tilknytning til forskellige discipliner. Eksempelvis kan alkoholisme behandles såvel i psykiatri, i psykologi, i sociologi m.v. Nogle brugere anvender som et selektionskriterium, at de kun eller fortrinsvis læser litteratur eller benytter håndbøger og bibliografier der eksplicit omhandler eet bestemt fag (som regel det, de er uddannet i eller under uddannelse til). Ofte er der imidlertid tale om, at litteratur, der eksplicit tilhører en bestemt videnskabelig disciplin også indeholder andre teorier, metoder etc. end litteratur fra en anden disciplin - og dem går man i så fald glip af. Det er naturligvis et helt andet problem, om man i sin undersøgelse af et bestemt emne ønsker

at anlægge en bestemt synsvinkel, f.eks. en psykoanalytisk, en fænomenologisk eller en materialistisk synsvinkel. Hvad jeg her taler om er, at værker ofte indeholder samme slags artikler, selvom de bærer titler, der indikerer, at de hører til forskellige fagområder. Hvis man f.eks. sammenligner indholdet i håndbøger i henholdsvis klinisk psykologi og psykiatri (tilhørende henholdsvis den psykologiske og den medicinske videnskabsorganisation) indeholder de ofte helt samme slags artikler, og forskellene er ofte mere et udtryk for redaktionens særlige kendskab og interesse end for egentlige disciplinære forskelle. Der vil derfor være tale om en uhensigtsmæssig brugerholdning, hvis man ser bort fra et relevant kapitel om et givent emne, fordi dette er trykt i en psykiatrisk håndbog, og man selv er psykolog. Man kan altså sige, at der kan være tale om en tendens til at f.eks. psykologer i højere grad læser/citerer andre psykologer, sociologer i højere grad læser/citerer andre sociologer o.s.v. - udover hvad emnet rent sagligt berettiger til i forhold til den eksisterende litteratur. Garfield (1979, p. 198) kan på baggrund af bibliometriske analyser bekræfte denne antagelse.

"Dette virker på mig som om psykologi er karakteriseret ved en høj grad af provinsionalitet. Jeg vil ikke gå så vidt som til at sige, at psykologer og adfærdsforskere ikke er særligt åbne overfor påvirkninger udefra, men de synes ikke at have meget at gøre med resten af forskningsverdenen. Hvis de har, indikerer litteraturen, at de ikke har fundet meget af værdi udenfor deres verden, eller hvis de har, vil de ikke indrømme det i deres referencer".

Psykologihistorisk er Garfields konklusion forkert. Psykologien er i stort omfang influeret af evolutionsteorien, statistikken, informationsteorien - og næsten enhver filosofisk grundantagelse har sit modstykke i en psykologisk teori. De af Garfield undersøgte artikler fra 1969 (hovedsageligt amerikanske) udviser ikke en sådan åbenhed, hvilket måske kan forklares ved forhold omkring amerikansk psykologis paradigme og forskningskultur, herunder det

forhold, at det måske er lettere at få accepteret artikler i de videnskabelige tidsskrifter, hvis disse refererer til andre psykologiske forskere, idet det dermed er vanskeligere at tilbagevise dem som uvedkommende eller perifere for faget.

Uanset forklaringen, er det her hensigten at konstatere, at disciplinært tilhørsforhold tilsyneladende er et anvendt selektionskriterium i litteratursøgning. Det er vel nok eet af de mest uheldige kriterier, der kan anvendes, idet de vil give en ensidig opfattelse af et problem, hindre tværdisciplinært samarbejde og nedsætte værdien af forskningen. Ganske vist kan der opnås en kortsigtet reduktion af overlæsningen, men hvis man vil have en dybere forståelse af emnet, må man før eller siden have fat på denne information, hvorfor den langsigtede virkning bliver en større informationsoverlæsning, fordi søgningen er grebet urationelt an fra starten.

Hvis man tager et emne som videnskabsteori for psykologer, vil de fremstillinger, der er skrevet eksplicit for psykologer ofte udgøre andenhåndsoplysninger i forhold til videnskabsteoriens generelle grundlitteratur eller til litteratur om samfundsvidenskabernes videnskabsteori. Man vil ofte køre længe rundt i denne sekundære litteratur uden at få en tilstrækkelig forståelse, hvorimod en direkte satsning på hovedværkerne selv ofte giver en hurtigere afklaring.

AFGRÆNSNING I TID.

En bruger kan have den holdning, at den nyeste litteratur er den mest relevante, og derfor begrænse sin litteratursøgning tidsmæssigt. Bag denne strategi kan ligge nogle helt forskellige antagelser, hvoraf den mest almindelige går på den videnskabelige litteraturs kumulative karakter, hvorved forstås det forhold, at den nyere litteratur optager og indarbejder tidligere litteraturs resultater. Denne antagelse har naturligvis relativ sandhedsværdi, større i nogle fag end i andre og for nogle problem-

stillinger end andre. I det omfang, den videnskabelige litteratur har en sådan kumulativ karakter er metoden effektiv, men det er min opfattelse, at den kumulative karakter af forskningen ofte overvurderes.

Der kan også bag den tidsmæssige begrænsning ligge den antagelse, at litteratur (og de heri indeholdte data) forældes, og derfor ikke har interesse, såfremt de er af ældre dato. Dette er naturligvis tilfældet på områder, hvor der har fundet en meget klar teknisk eller videnskabelig udvikling sted, således at de observationsmuligheder, der var til rådighed for tidligere tiders videnskabsmænd klart er forældede. Imidlertid er det ofte forbløffende, hvor præcise og relevante tidlige observationer har været - måske fordi tidligere tiders forskere havde mere tid til at hellige sig videnskaben.

Den tidsmæssige begrænsning kan også hænge sammen med opfattelsen af en problemstillings oprindelsestidspunkt eller forskningsgenstandens eksistensperiode. Det er naturligvis forholdsvis let i de teknisk-naturvidenskabelige fag at afgrænse f.eks. litteratur om transistorer og andre konkrete opfindelser, og derved sætte en effektiv baggrænse for litteratursøgningen. I human- og samfundsvidenskaberne kan man sige, at forskningslitteraturen ofte følger samfundsmæssige fænomener. Således er f.eks. litteratur om arbejdsløshedens psykologiske følger skrevet i perioder med arbejdsløshed og man kan f.eks. afgrænse litteratursøgningen bagud til 1930'erne. Ofte er der imidlertid hos litteratursøgere nogle helt urealistiske forestillinger om, at et emne er af forholdsvis ny oprindelse, hvorfor man sparer sig selv for at søge ældre litteratur. En sådan manglende historisk bevidsthed fører til gentagelser af tidligere forskning og til en overfladisk opfattelse af genstandsområdet. (I en artikel (Hjørland, 1984b) diskuterer jeg, hvordan moderne edb-baseret litteratursøgningsteknik muligvis kan forstærke tendensen til en manglende historisk fundering af forskningen).

Tidsmæssig begrænsning af litteratursøgningen må derfor siges at være en effektiv metode til nedbringelse af informationsoverlæsning, hvis ovenstående forhold er taget i betragtning (hvilket de som regel langt fra er). For at opnå nogle af metodens fordele og samtidig sikre en større historisk dimension, kunne man måske tænke sig en kompromismetode, der bestod i en springvis søgning, hvor man foruden det nyeste tog stikprøver med et regelmæssigt tidsinterval så langt de bibliografiske hjælpemidler rækker.

(Den tidsmæssige begrænsning giver brugeren en sekundærgevinst. Ved at referere til den moderne litteratur viser han indirekte, at det emne han beskæftiger sig med er "oppe i tiden", at det er et emne, som også andre finder relevant og moderne - og at problemstillingen eller synspunkterne ikke er forladt.)

AFGRÆNSNING PÅ SPROG, FORM, Udstyr, Sted og Tilgængelighed.

Sprogbarriererne (overfor fremmedsprog) spiller en indlysende og selvstændig rolle som et afgrænsningskriterium i videnskabelig litteratursøgning. Såvel de herskende normer m.h.t. hvad forskere må beherske som konsekvenserne af sprogbarriererne er forskellige fra fag til fag. I medicin er engelsk næsten blevet helt accepteret som standardsprog, og denne forsknings karakter gør konsekvenser af manglende sprogfærdighed mindre end i f.eks. samfundsvidenskaberne, hvor sproglige forskelle ofte også dækker over ganske betydelige substantielle forskelle i metode, teori og fænomen. Generelt sagt må man sige, at sprogbarrierer ikke bør kunne forhindre, at værdifulde forskningsresultater opsuges i den efterfølgende forskning. Ansvar for problemer af denne art kan ikke kun placeres hos den enkelte forsker, men er i høj grad et kollektivt ansvar bl.a. om effektive oversættelsesservicer og referatidsskrifter. Netop i en situation, hvor forskeren overlæsses med videnskabelig information, kan fristelserne til at afgrænse søgningen til de lettest tilgængelige sprog være særlig stor. Først hvis forskeren får fornemmelsen af,

at samme emne er behandlet kvalitativt anderledes på andre sprog vokser motivationen til at studere denne litteratur. Det kan derfor antages, at generelle artikler om forskningsapproach i forskellige lande kan være værdifulde, ligesom det kan være værdifuldt at få resuméer af litteratur, selvom man ikke behersker originalsproget. For en mere detaljeret beskrivelse af sprogbarriererne kan henvises til Large (1983).

Et helt andet sprogligt problem er, hvad man kan kalde fagsproget eller det videnskabelige sprog. Der er vigtige forskelle i måden at anvende sproget på i videnskabelig litteratur - både i primærlitteraturen, d.v.s. tidsskrifterne og bøgerne og i det, der under eet kaldes dokumentrepræsentationer, d.v.s. titler, emneord, referater o.s.v., som i første omgang er afgørende for litteratursøgningens afgrænsning. Disse sproglige forskelle kan have stor - positiv eller negativ - betydning for overlæsningen.

Der kan naturligvis være mere eller mindre gode grunde til, at videnskabelige dokumenter eller deres repræsentationer får den ene eller den anden sproglige udformning. Det interessante i forbindelse med vor problemstilling er de faktiske konsekvenser, som en bestemt sproglig udformning kan have for litteratursøgere m.h.t. beslutningsgrundlaget for en nærmere undersøgelse af givne dokumenter. Med andre ord: hvorvidt en given sproglig udformning enten kan være skyld i at iøvrigt relevante dokumenter vurderes som irrelevante eller modsat: at forskere overlæsses, idet mange dokumenter p.gr.a. sproget i første omgang ser lovende ud, hvorfor forskeren spilder tid på at fremskaffe og læse dokumenterne, blot for at konstatere, at de alligevel ikke er relevante, d.v.s. de sproglige forhold forårsager overlæsning.

I et tidligere afsnit har vi set på, hvordan litteratursøgere kan anvende disciplinære tilhørsforhold som kriterier for afgrænsning og relevansvurdering. Vi kan nu give nogle eksempler på, hvordan sproget kan anvendes til at definere den disciplinære identitet. Det forholder sig her, som det iøvrigt kendes

fra socialpsykologien: forskellige grupper anvender forskellige "signaler" hvormed man kan fastslå, om man er "in" eller "out" i en given social sammenhæng (f.eks. signaler som frisure, klædedragt, sprog o.s.v.). I forretningskredse er der således knyttet positive associationer til begrebet "dynamiske personer", medens der er knyttet negative associationer til samme begreb i pædagogkredse, hvor man til gengæld gerne må være "gruppedynamisk".

Indenfor litteraturen er der en tilbøjelighed til at marxistiske forskere efterligner Marx' terminologi og kalder deres artikler "Til kritikken af X" i stedet for blot "Om X". Omvendt er der i den positivistiske litteratur en tilbøjelighed til at efterligne naturvidenskabeligt sprog. Således er der ikke kvinder i henholdsvis den største pædagogiske og den største psykologiske thesaurus, men "human females", d.v.s. en klar biologisering af terminologien. Disse to eksempler viser, hvordan forskellige retninger ikke blot kommunikerer deres argumenter rent sagligt, men også benytter en terminologi, der identificerer den pågældende retning alene på sprogbrugen. Hermed er man med til at fremkalde "in-gruppe" eller "out-gruppe" fornemmelser hos læseren, og i sidstnævnte tilfælde et negativt førstehåndsindtryk og dermed en mindre chance for, at dokumentet overhovedet bliver læst. Sproget kan altså blive en ekstra barriere for videnskabelig kommunikation, ligesom det er med til at opretholde de kløfter mellem discipliner, som tidligere er diskuteret.

I naturvidenskabelige kredse eller de kredse, der gerne vil ligne naturvidenskaberne, f.eks. adfærdsvidenskaberne, kan der være en forkærlighed for matematiserede udtryk eller latiniserede betegnelser udover disses saglige berettigelse. Der kan også være tale om meget ordknappe fremstillinger, der oprindeligt har haft deres berettigelse i tidsskrifts- eller læser-

økonomiske forhold, men som efterhånden er selvstændiggjort som norm for hvordan videnskabelige dokumenter bør udformes. (Herunder hører f.eks. tendensen til at forkorte forfatteres fornavne til bogstaver, idet kvinders fornavne dog efter lidt ældre normer ikke forkortes). Denne slags sprogbrug kan være med til at give indtryk af større videnskabelighed end der nærmere set er dækning for, og dermed til at fremkalde falske forestillinger om relevans og deraf følgende overlæsning. Der er tale om, at videnskabelige dokumenter skal give området prestige, og derved ofte dække over utilstrækkelig viden. Nogle af de her omtalte sproglige mekanismer er velkendte i fagkredse og har været genstand for vittige kommentarer. Som illustration hertil bringes nedenfor et udtog af nogle eksempler på "S-L-Koden" (Scientific Language Code), der har været bragt som nytårshilsen i tidsskriftet "Skolepsykologi" i 1977.

S-L-Koden

(Scientific Language Code)

Koder har de seneste årtier interesseret den pædagogisk-psykologiske verden.

Bernstein gav startskuddet til denne spændende forskning, som nu har spredt sig til mange lag, og det er blevet en videnskab værdigt at løse koder. Efterretningstjenester og drengeblade har ikke længere patent på dette.

Vi kender problemer med kodning af de budskaber, der er indeholdt i udtalelser fra tidligere arbejdsgivere

socielt indstillet = *skørtejæger / for lange kaffepauser*

fra skolers udtalelser om elever

viste ofte interesse = *mødte sjældent*

eller

fremtidsorienteret = *har længe ventet på at blive udmeldt*

Med bare lidt træning vil man forstå at tyde indholdet for palydende.

Også den videnskabelige faglitteratur har i årtier haft sin kode, der for lægløkk og naive kolleger har været uigennemskuelig.

SKOLEPSYKOLOGI har valgt at afslutte dette års publikationer med en forbrugervenlig guide, der med ét bortdriver det kunstige tagesløb, der hidtil har indhyldet mange videnskabelige artikler. For at hjælpe ved læsernes fremtidige aktive læsning præsenterer vi her de væsentligste sløringsmetoder og den korrekte oversættelse

Man har længe været bekendt
med, at ...
... jeg gad ikke slå originalkilden op

.. af stor teoretisk og praktisk
betydning ...
... interesserer mig

Q

Q

.. omend det næppe er muligt på
helt entydig vis for nærværende, at
give endegyldige svar på disse
spørgsmål ...
... der er overhovedet ikke kommet
noget ud af forsøget,
men jeg synes nu alligevel, at
jeg ville udgive én publikation

.. åbenbart, selvfølgelig ...
... andre har også allerede
skrevet det

Q

subjektivt ...
... de andres mening

Q

Q

Utvivlsomt forestår der endnu
mange og dybtgående
undersøgelser,
før man kan nå frem til en
fuldstændig forståelse ...
... jeg forstår det ikke.

Bedrøveligvis er der endnu ikke
formuleret en kvantitativ teori,
der viser denne effekt ...
... der er ingen, der ved
nogetsomhelst om det.

Q

Q

Det er vort håb, at dette arbejde
må inspirere til
nærmere undersøgelse ...
... undersøgelsen er faktisk
temmelig slet,
men de andres udgydelser
er i hvert fald lige så dårlige.

objektivt ...
... min mening

Q

det er almindeligt antaget, at ...
... jeg og et par andre mener

Q

Q

.. det er velkendt, at ...
... jeg har tilfældigvis hørt

Resultaternes overensstemmelse
med den teoretiske kurve er:
overordentlig stor ...
... nogenlunde
god ...
... næppe til at se
antagelig ...
... overhovedet ikke til stede

Q

Q

Tre forsøgspersoners resultater
blev udvalgt til nærmere analyse ...
... resultaterne fra de øvrige gav
ingen mening og blev lagt væk

.. tendentiøs retning ...
... fusk

Den videnskabelige litteratur underlægges i stadig større udstrækning markeds kræfter og i stadig mindre udstrækning alene videnskabelige samfunds- og institutioners kvalitative evalueringsskriterier. Dels er der som tidligere nævnt tale om, at stadig flere videnskabelige tidsskrifter udgives af kommercielle forlag fremfor non-profit organisationer (som iøvrigt også bliver mere og mere "økonomisk bevidste"), dels er der tale om, at forskere og forskningsinstitutter ser en stadig større interesse i at deres publikationer citeres - fordi dette kan måles i citationsindex og dermed have mulige konsekvenser for bevillinger m.v. Det vil igen sige, at forskere kan være fristet til - blandt andre metoder - at anvende en reklameterminologi for at få øget afsætning eller benyttelse af deres publikationer. Mange værker kalder sig nu "Håndbøger", men er i realiteten antologier, det samme gør sig gældende med serier "New advances in ...", hvor man ofte søger forgæves efter noget, der ligner egentlige fremskridt. Det at anvende adjektiver som "effektive", "objektive", "eksakte" o.s.v. i titler må også anses for reklameprog, idet det må være det naturlige krav til ethvert videnskabeligt arbejde, at det er så objektivt, effektivt og eksakt, som det pågældende problemområde nu engang lader sig behandle. Sådanne titler søger ofte at sælge sig selv på et frustreret behov hos forskere, hvis emneområde netop savner disse kvaliteter - uden dog at kunne leve op til det, som titlen lover.

Den mest benyttede - og skadelige - form for reklameterminologi er imidlertid nok at anvende bredere betegnelser, end der er dækning for i dokumentet. Det er naturligvis lettere at lave en videnskabelig undersøgelse af et ganske snævert emne end af et bredt, mens omvendt et bredt emne har en langt større benytterkreds. Derfor ligger der en stor fristelse i at generalisere dokumenters emne eller omvendt: undlade at konkretisere værkers specifikke emneområde.

Hvis dokumenter selv er tilbøjelige til enten reklameterminologi eller i al fald utilstrækkelig præcisering af emnet, så bør det

naturligvis være referattidsskrifters, informationssystemers m.fl.s opgave at råde bod på sagen og via kvalitativ indexering, systematik og referat at give den nødvendige præcisering af dokumenters emneindhold. I en vis udstrækning lever sådanne informationstjenester op til deres ansvar, men ofte er der tale om, at de - f.eks. på grund af manglende økonomi, tid eller kvalificeret personale - ikke foretager en tilstrækkelig selvstændig stillingtagen og beskrivelse af dokumenterne, men at personalet på disse tjenester selv til en vis grad bliver offer for dokumenternes reklamesprog. Hvis indexører arbejder på akkord, og hvis kvaliteten af deres arbejde bliver testet på graden af overensstemmelse med andre indexører, kan man let forestille sig, at dette ikke resulterer i en selvstændig stillingtagen til dokumenters emne, med derpå følgende oversættelse til systemets genfindingssprog. I stedet for vil indexørerne ofte gribe til den lettere udvej at vælge de thesaurustermene eller andre emnekoder, der ligger nærmest op ad f.eks. dokumentets titel eller indholdsfortegnelse, og altså forstærke virkningen af reklame-terminologi i stedet for at mindske den. Kvaliteten af indexering er naturligvis forskellig fra det ene informationssystem til det andet, så de her fremførte forhold gælder naturligvis ikke generelt. I almindelighed må vi imidlertid konstatere, at f.eks. referattidsskrifters indexører arbejder bedst med de begreber, der vedrører forholdsvis klare og håndgribelige kriterier, såsom dyrearter, aldersgrupper, kemiske forbindelser, tider og steder o.s.v. og dårligst, hvor det drejer sig om mere abstrakte eller kontroversielle begreber, f.eks. betegnelser for forskellige teorier og retninger. Dette har naturligvis en direkte sammenhæng med det forhold, at sådanne tjenester ofte kører med personale uden tilstrækkelig videnskabelig baggrund for at foretage den præcise emneanalyse. (Dette emne er undersøgt nærmere af Krarup og Boserup (1982)). Det kan tilføjes, at dette problem kan være ret så grelt, og at der decideret kan være tale om at referattidsskrifter tilfører støj, som i det tilfælde, hvor personalet i Psychological Abstracts sammenblander "gestalt psychology" med "gestalt therapy" (Jfr. Simkin, 1978).

Foruden det forhold, at de sekundære litteraturformer og informationssystemer har en potentiel mulighed for enten at forstærke en given "reklameeffekt" (hvis de fungerer for dårligt) eller at modvirke denne effekt (hvis de fungerer godt) har de andre effekter. Birgitte Lau & Karen Risager (1979) har f.eks. påpeget faren for, at disse systemers sprog bliver normdannende på en måde, der er med til at fortrænge modersmålet eller mere lokalt udviklede fagsprog. En sådan sproglig ensretning kan naturligvis være med til at øge overlæsningen, idet den mindsker mulighederne for at udtrykke mere nuancerede forhold - hvilket er særligt alvorligt i de humanistiske fag.

I det ovenstående har jeg forsøgt at vise, at dokumenttermer (især titelter og deskriptorer) kan fremvise en mere eller mindre bevidst eller tilsigtet reklamefunktion. I det omfang den videnskabeligt informations emballage og reklame bliver faktorer, der påvirker forbrugernes valg, d.v.s. i det omfang sproget ikke blot formidler "varefakta", men har held til at anprise sin vare, i det omfang vil brugeren spille sin tid med at fremskaffe og læse stof, der ellers ikke ville være læst, og hans overlæsning vil øges.

Strategien imod denne overlæsning er naturligvis først og fremmest et grundigt emnemæssigt overblik og sagkundskab, hvorved tom tale formindsker sin påvirkningsmulighed. Desuden vil det være gavnligt med størst mulig bevidsthed om sproglige og terminologiske forhold og med anvendelsen af andre selektionskriterier end de, der bygger på dokumenters sproglige udformning.

De sproglige forhold i forbindelse med udformningen af moderne edb-baserede søgesystemer rejser en helt anden type af overlæsningsproblemer. De fleste systemer satser mest på fritekstsøgning, hvor enkeltord i søgningen skal kombineres til en søgestrategi. Et emne som "hvordan informerer man bedst børn om forældres hospitalsindlæggelse" vil typisk skulle søges ved at

kombinere ordene "parents", "children" og "hospitalization". Sådan ville man ikke gøre i gammeldags, hånddrevne, "prækoordinerede" registre, der ville man typisk have et opslagsord for "hospitalization" og en underterm for "child hospitalization" (der udgør langt den største gruppe). Hvis der ikke var indførsler nok til, at det er rimeligt at give "adult hospitalization" sin egen indførsel, ville man kunne søge dette enten under "hospitalization" i almindelighed eller som en opsamlingsgruppe "andre former for hospitalisering". Denne gammeldags registertype kunne altså løse det problem at lokalisere litteraturen om voksnes hospitalisering uden samtidig at skulle gemme den meget store kategori af dokumenter, der omhandler børns hospitalisering. Det kan edb-systemet også klare ved at kombinere "adults" med "hospitalization", men når man så får ovenstående emne stukket ud, er man på den, fordi man jo også har brug for at definere, at emnet vedrører børn. Der er ikke større principielle vanskeligheder i at løse problemet. Man kunne have løst problemet ved at have en term for "child hospitalization" og en for "adult hospitalization", men det har man ikke, og nogle gange møder man ligefrem folk, der mener de har ekspertise på dette felt, og som alligevel fraråder brugen af flerordsdeskriptorer - uden at de kan angive andre løsninger på de mange "false drops" og den overlæsning, dette medfører (f.eks. Ørnager, 1983).

Dette og andre sproglige problemer i forbindelse med edb-søgning kan give helt urimelige "overload". Det er meget varierende, hvor effektive man kan gøre sine søgninger, men man skal være glad, hvis støjen er under 50%. I et grelt tilfælde dokumenterer Brand og Kinzie (1984), at man ved søgning på emnet "rekonstruktion i psykoanalyse af børn" kun finder tre % relevante dokumenter i en af de mest oplagte databaser. D.v.s. hvis man skal have dokumenter nok til en artikel eller et speciale skal man betale for og bruge tid på at gennemgå 100 dokumenter, hver gang man får tre relevante. For at opnå en rimelig dækning skal man altså gennemgå mange hundrede, måske 1.000 dokumenter. Det kan i sandhed kaldes overlæsning. Problemer af denne art forstærkes, efterhånden som databaserne påfyldes flere og flere

referencer, men problemerne er løselige, blot kræves der en højere grad af ekspertise både ved konstruktionen af søgesprogene og ved indexeringen af det enkelte dokument. I begejstringsrusen over edb-søgningens effektivitet er disse problemer blevet negligeret, men netop den rationalisering, der ligger i edb-teknikken bør betyde bedre økonomiske muligheder for kvalitetsforbedringer.

Jeg vil slutte gennemgangen af de fagsproglige problemer ved at foretage en distinktion mellem den terminologi, der direkte vedrører en given forskningsgenstand (dens egenskaber, processer o.s.v.) og den terminologi, der - metavidenskabeligt - beskæftiger sig med, hvordan man tidligere har udforsket denne genstand, d.v.s. termer for metode, teori, retninger etc. De moderne edb-baserede informationssystemer sigter især mod det første, mod at den uforberedte bruger kan stille sit spørgsmål og få sit svar. Det er en naiv-realistisk antagelse at tro, at hver litteratursøger kan udforske verden uden at kende indholdet af de videnskabelige retninger og skoler, der har forsøgt at udforske denne med forskelligt held. Et sådant approach øger overlæsningen på længere sigt, fordi den overlader det til brugeren selv at gøre de erfaringer, som er ophobet i videnskabens historie. Derfor udgør en sådan "metaterminologi" et forsømt område, ligesom sammenhængen mellem "realterminologi" og "metaterminologi" gerne skulle afspejles i f.eks. thesauri. Indenfor psykologien er masochisme f.eks. et helt overvejende psykoanalytisk begreb, der ikke anvendes af behaviorister, der kun tangerer det i begrebet "self-punishment". Dette oplyser ingen psykologisk thesaurus, og det ville være nok så nyttigt i litteratursøgning.

Litteraturformer og -udstyr.

Brugere kan begrænse deres litteratursøgning ved hovedsageligt at satse på visse litteraturtyper fremfor andre. Såfremt motiveringen herfor er, at nogle publikationsformer er mere fremelskede af markedsmekanismerne, så hører det ind under de tidligere behandlede indirekte indikatorer for værks eller forfatters kvalitet. Der kan imidlertid være andre motiver. Dels

kan man forestille sig en - vel nok temmelig ubevidst - konservatisme, der især indskrænker læsningen til den slags materiale, man er vant til. Overbibliotekar Torben Nielsen hævdede engang med beklagelse, at studerende nu om dage kun er vant til at læse uddelte duplikater, og bliver helt forvirrede, hvis man foreslår en bog. Personligt er jeg nu nok mere tilbøjelig til at mene, at mange forskere kan have en vis konservatisme i netop at holde sig til bøger og underudnytter tidsskrifterne. Et andet alternativ er, at nogle brugere indskrænker deres læsning til et enkelt medium på grund af særlige forventninger til dette stadium, f.eks. til tidsskriftsartikler p.gr.a. særlige forventninger om det videnskabelige niveau grundet redaktionsprocessen.

Endelig for det tredje er det et kendt forhold, at studerende ved læreanstalterne med forkærlighed læser hinandens opgaver. Herved opnår de - eller mener de at opnå - foruden den informationsværdi, der modsvares af en tilsvarende bog eller mængde artikler - indirekte nogle oplysninger om, hvordan en opgave struktureres, om kravniveau etc. I og med at dette er tilfældet for studenterarbejder, skal der ikke megen fantasi til at forestille sig, at det samme - i mindre udstrækning - gælder også for disputatser, for artikler o.s.v. (men naturligvis aftagende, jo mere erfaring forskeren får med medierne). Vi kan således formulere den hypotese, at en forsker - alt andet lige - vil gøre lidt større brug af den litteraturform, som han selv stiler mod at publicere i, i forhold til, hvordan litteraturen rent faktisk fordeler sig på litteraturformerne.

Denne hypotese kan belyses med bibliometriske data fra Line (1981, p. 84), der viser, at i samfundsvidenskaberne som helhed citerer bøger 39% tidsskrifter, 51% bøger og 10% andet, medens tidsskrifter citerer 47% tidsskrifter, 25% bøger og 28% andet. Det forholder sig altså faktisk sådant, at bøger især citerer bøger og tidsskrifter især citerer tidsskrifter. Vi kan imidlertid ikke vide, i hvor høj grad dette skyldes den omtalte psykologiske tilbøjelighed hos forfatterne og i hvor høj grad det skyldes, at nogle delområder (f.eks. eksperimentopsykologi) især er repræsenteret i tidsskriftslitteraturen, der derfor citerer sig selv, mens andre emner (f.eks. psy-

kologihistorie) i højere grad er repræsenteret i monografilitteraturen, der derfor også må have en overvægt af referencer til monografier. Vi kan således ikke underbygge vor hypotese, men blot fastslå, at foreliggende data sandsynliggør den og ikke strider imod den.

Indenfor samfundsvidenskaber og psykologi er der en tilbøjelighed til, at referattidsskrifter og informationstjenester overvurderer tidsskrifternes betydning på bøgernes bekostning. Dette er beklageligt, og i grunden burde det være overflødigt at argumentere for, at det også er vigtigt at dække monografilitteraturen. I en brevveksling med redaktionen af Psychological Abstracts om den manglende dækning af monografier, har man svaret forfatteren til denne artikel, at dette måtte være et europæisk problem, fordi amerikansk psykologisk forskning kunne baseres på artikler alene (sic!). I og med dette synspunkt altså kan fremsættes, mener jeg også, det må være på sin plads her at tilbagevise det. Jeg vil fremføre to eksempler på psykologiske forfattere, der mener, at monografilitteraturen tilfører noget særligt og væsentligt i forhold til tidsskriftslitteraturen. Westland (1978, p. 95) beskriver, at kritikken af den psykologiske forskningspraksis næsten aldrig ses i de egentlige forskningstidsskrifter, men fremgår af uformelle kontakter, af bøger, af andre typer tidsskrifter (fagblade) og tidsskrifter fra nabo-discipliner. Dette var den første dokumentation for bøgers berettigelse. Den anden leveres af en anmelder i Contemporary Psychology (1978, 23, p. 803-806), der kritiserer en håndbog i arbejdspsykologi og mener, at en del af forklaringen skal søges i, at nogle - selv eminente - psykologer indskrænker deres læsning til bidrag, der har været trykt i tidsskrifter, mens hovedværker, der har bidraget til områdets udvikling er groft underrepræsenteret. Man spørger sig selv, om disse redaktører og forfattere i for høj udstrækning har baseret deres litteratursøgning på Psychological Abstracts? Disse to eksempler er naturligvis langt fra de eneste, der kan fremføres, men som sagt finder jeg sagen så banal, at det slet ikke burde være nødvendigt at dokumentere.

Et psykologisk speciale, udgivet af Psykologisk Institut i Århus (Zachariae, 1983) satser næsten udelukkende på monografier, idet kun tre ud af 133 referencer er til tidsskriftsartikler. Dette er noget påfaldende. En del af forklaringen ligger vel nok i, at en stor del af den benyttede litteratur er af "antipsykiatrisk" orientering - og netop en sådan orientering er næsten defintorisk underrepræsenteret i de egentlige forskningstidsskrifter og dermed databaser. Desuden er der nok også tale om, at forfatteren til dette speciale har anvendt en litteratursøgningsstrategi, der helt baserer sig på bøger (fra bibliotekernes åbne hylder? kartotek? egne bogkøb?), en strategi, der må siges at være heldig i dette tilfælde, idet udgiverne er fulde af lovord i forordet. En sådan strategi medfører en stor del hovedværker (f.eks. Freud, Marx, antipsykiaterne) og grundbøger (f.eks. videnskabshistoriske fremstillinger), og må nok siges at være relevant netop i specialesituationer og i situationer, hvor utilfredsheden med det/de eksisterende paradigme(r) tvinger forskeren til en bred ny-orientering. Derimod kan en mere rutinepræget forskning næppe på samme måde tillade sig at se bort fra tidsskrifterne. Her anses grundbøger mere for et passeret stadium, og forskningsprocessen består nærmere af en sammenvævning af højt specialiserede bidrag på et trindhøjere niveau. Hvis vi betragter det at låne bøger på biblioteket som et første trin, det at udnytte disse bøgers referencer som det andet trin, det at udnytte andettrinnets referencer som et tredje trin o.s.v. kan man opfatte forskningsprocessen som en flertrinsraket. Det er klart, at specialer og tidsbegrænsede projekter ikke på samme måde som professionel forskning kan blive en flertrinsraket, men man må til gengæld formode, at jo flere trin der gennemløbes, desto større vil sandsynligheden blive for at referencerne kommer til at afspejle litteraturens sande fordeling på publikationsformer, idet chancen for at overse relevante artikler og andre former mindskes jo flere kilder, der inddrages.

Foruden forholdet mellem artikler, bøger og mimeograferede arbejder, gør også forholdet mellem primære og sekundære litteraturformer sig gældende, d.v.s. forholdet mellem bøger/artikler

på den ene side og f.eks. håndbøger på den anden. Et håndbogs-kapitel vil ofte minde noget om de tidligere omtalte "literature reviews", og hvad der blev sagt om dem, gælder også her: gode håndbogsoversigter nedbringer overlæsningen, men dårlige og ukompetente oversigter øger den. Derfor kan bibliotekarere ikke blot benytte en stående tommelfingerregel om at anviser brugere håndbøger, for herved at begrænse deres overlæsning, således som Weicks opfattelse (1970) implicerer. Der er heller ikke på dette plan nogen vej uden om problemet med at tage kvalitative hensyn.

Ofte kan håndbøgers størrelse og udstyr indgyde en autoritet, der kan gøre brugere (og intermediære, f.eks. bibliotekarere og universitetslærere) tilbøjelige til at satse på dem. Det samme kan udstyr, f.eks. en fyldig referenceliste. Tidligere var f.eks. en fyldig referenceliste også i højere grad end i dag en garant for faglig kvalitet, men også sådanne midler anvendes i dag som metoder til at imponere læseren. En anmelder i *Contemporary Psychology* (Rosenbaum, 1982, 350-352) konstaterer således, at sammen med krisen i socialpsykologiens indhold, svulmer antallet af referencer i lærebøgerne mere og mere op. Man fornemmer, siger han bl.a., at det store antal referencer dækker over en vis usikkerhed omkring det adekvate i de foretagne konstateringer. Ofte henviser referencerne kun til banaliteter, og der forsøges henvist til alle tænkelige empiriske og teoretiske pointer, af hvilke der er mange. En sådan citeringspraksis finder man efter Rosenbaums mening ikke indenfor naturvidenskabernes lærebøger eller indenfor mere sunde områder af psykologien.

Som strategi til at nedbringe overlæsning må satsning på bestemte litteraturformer i almindelighed anses for u hensigtsmæssig, fordi det vil være helt arbitrært, hvad der udelades. Skal der foretages et valg, så bør bøger nok have højest prioritet, idet disse har mest historisk kontinuitet, videnskabsteoretisk spredning, integration af anden litteratur m.v. En undtagelse udgør

de tilfælde, hvor man er interesseret i rent eksperimentelle resultater, hvor tidsskrifterne er det vigtigste medium.

Litteratur fra bestemte lande/steder.

Hvis vi ser bort fra de tidligere behandlede fremmedsprogsbarrierer, kan mange brugere være tilbøjelige til at anvende et geografisk kriterium i litteratursøgningen, d.v.s. at foretrække f.eks. sovjetisk eller amerikansk litteratur om et emne. I samfundsvidenskaberne kan sådanne præferencer naturligvis høre sammen med samfundsforholdenes indflydelse på forskningsgenstanden, d.v.s. at bestemte emner kun optræder i bestemte geografiske eller sociale sammenhænge eller får særlige kvaliteter i sådanne sammenhænge. Der kan imidlertid også være tale om en mere generel tillid/mistillid til bestemte områders forskningsmæssige værdi (hvilket kan hænge sammen med uddannelsessystemet, forskningstraditioner og forskningsvilkår).

Iøvrigt er der vel nok alt andet lige sandsynlighed for, at forskere prioriterer litteratur fra deres eget land højest (hvilket f.eks. kan spores i tydeligt nationale forskningstraditioner på mange forskningsområder).

Det geografiske kriterium kan være hensigtsmæssigt til nedbringelse af overlæsning, men må siges at være et groft kriterium. Det vil ofte forudsætte viden om forskellige landes forskningstraditioner, og såvel sprogbarrierer som det bibliografiske apparats ufuldstændighed vil ofte sætte en grænse for, hvor langt dette kriterium lader sig udnytte.

Fysisk tilgængelighed.

Det er en veldokumenteret kendsgerning, at forskere foretrækker litteratur, der er "ved hånden" fremfor litteratur, der er svære at fremskaffe. Enten må forskerne hermed gå ud fra, at det ikke spiller nogen større betydning for deres arbejde hvilken litteratur de benytter (f.eks. fordi de forestiller sig en betydelig redundans, så bøgerne "ved hånden" i realiteten redegør

for de samme forhold som de vanskeligt tilgængelige) eller også er deres forskningsmotivation således, at de foretrækker at komme let til dårlig forskning fremfor at have mas med en mere værdifuld forskning. Gerstberger og Allen (1968) har vist, at ingeniører tilsyneladende foretrækker den lettest tilgængelige kilde, uanset deres opfattelse af dens relative værdi i forhold til andre informationskilder. Undersøgelser af denne art kan give en fornemmelse af, at megen moderne forskning ikke satser hårdt på egentlige erkendelsesfremskridt, men stiller sig tilfreds med den pseudoaktivitet, det er at skrive artikler, der opfylder forskersamfundets formelle kvalifikationskrav. Herved kommer vi ind i en ond cirkel, thi det nedsætter informationsværdien af den videnskabelige litteratur, og gør det ekstra tillokkende at undgå besværligt biblioteks- og litteraturarbejde.

En nedbringelse af informationsoverlæsningen ved at satse på let tilgængelige kilder kan være nødvendigt i konkrete situationer og i meget tidsbegrænsede forskningsprojekter, f.eks. semesteropgaver på universiteterne, men må i almindelighed anses for ganske uvidenskabelig og uhyre kortsigtet.

KONKLUSION

I denne artikel har jeg forsøgt at give en systematisk og afrundet oversigt over de vigtigste selektionsstrategier, der står til en forskers rådighed, når han overlæses med videnskabelig information. Det er et af mine hovedsynspunkter, at problemstillingen ikke kan "psykologiseres", d.v.s. det er ikke i første omgang forskerens manglende kendskab til rationel litteratursøgning, der forårsager problemet, men hans kollegers og offentlige myndigheders utilstrækkelige indsatser, der er den primære faktor. Jo mere "stresset" et forskningsområde bliver, jo mere kortsigtet forskere tvinges til at arbejde, desto større bliver problemerne for andre forskere med at skille de værdifulde bidrag ud fra trivialiteterne. Udforskningen af, hvordan sådanne trivialiteter produceres kan siges at være et videnskabssociologisk eller videnssociologisk problem. Patrick Wilson er en

af de meget få forfattere, der har taget disse problemer op decideret ud fra en biblioteks- og informationsvidenskabelig synsvinkel. Samtidig med udforskningen af informationens produktion er det også interessant, hvordan den enkelte forsker bærer sig ad med at orientere sig i litteraturen, i hvilken grad han gennemskuer de tilsløringsmekanismer, der udlægges, kort sagt hans adfærd, viden og holdning i litteratursøgningsarbejdet. Dette er den mere psykologiske del af informationsvidenskaben, traditionelt betegnet som "brugerundersøgelser". Det er altså eet af mine ønsker, at disse brugerundersøgelser må blive mere præget af en udforskning af, hvordan litteraturen reelt er beskaifen og produceres. Når dette er gjort, er brugerundersøgelser naturligvis værdifulde for biblioteks- og informationssektoren, fordi det er dennes opgave at højne informationssøgningens standard, og dette gøres bedst ved at medinddrage brugernes psykologi.

Næsten ethvert punkt i min artikel kan danne udgangspunkt for yderligere forskning. Som sagt i indledningen ville eksempelvis undersøgelser af forskeres brug af anmeldelser og "reviews" være værdifulde. Sådanne punktvisse belysninger fremkommer da heldigvis også lejlighedsvis i litteraturen. Foruden punktvisse undersøgelser vil det imidlertid være særdeles ønskeligt med mere sammenhængende undersøgelser i et eller flere forskningsområder. Det ville være ønskeligt at vide noget om, hvordan alle de selektionsmuligheder, der er omtalt i denne artikel rent faktisk udnyttes i praksis, f.eks. hvor stor en andel af litteraturen i et givet forskningsprojekt, der selekteres på baggrund af direkte check af primærdokumentet, hvor stor en del af forhåndskendskab til forfatteren o.s.v. En sådan undersøgelse vil kræve, at man på forhånd har en nogenlunde systematisk og afrundet oversigt over "udfaldsrummet", idet man i modsat fald løb en risiko for, at såvel undersøgeren som brugeren glemte vigtige enkeltfaktorer, og dermed fik et forkert helhedsindtryk. Det er forbavsende, at en så forholdsvis elementær ting ikke tidligere er udarbejdet.

Patrick Wilson (op.cit.) redegør for, at baggrunden for hans bog er et ønske om at bibliotekarere skal fungere i informationssystemer

og ikke i misinformationssystemer: at brugeren skal informeres og ikke misinformeres. Jeg vil slutte med at gøre disse ord til mine egne.

Indholdsfortegnelse:

INDLEDNING	F1
FAKTORERNES SYSTEMATISERING	F2
DIREKTE CHECK AF PRIMÆRDOKUMENTER	F4
DIREKTE KENDSKAB TIL FORFATTEREN	F5
INDIREKTE KENDSKAB TIL VÆRK ELLER FORFATTER	F7
Mest citerede forfattere	F7
Mest publicerede/efterspurgte bøger	F9
Andre indirekte indikatorer for værks/forfatters kvalitet	F10
KENDSKAB TIL FORSKNINGSOMRÅDET	F11
Mest udbredte emner og synspunkter	F11
Litteratur fra bestemte discipliner	F12
AFGRÆNSNING I TID	F14
AFGRÆNSNING PÅ SPROG, FORM, Udstyr, Sted og Tilgængelighed	F16
Sprogbarrierer (fremmedsprog)	F16
Fagsproglig udformning	F17
Litteraturformer og udstyr	F25
Litteratur fra bestemte lande/steder	F30
Fysisk tilgængelighed	F30
KONKLUSION	F31
INDHOLDSFORTEGNELSE	F33
REFERENCER	F34

REFERENCER.

Brand, Alice A. eg L.A.Kinzie: A comparison of online access to psychoanalytic literature. Database, 1984, 54-63.

Clod Poulsen, Sten: Litteraturstudier som led i pædagogisk-psykologisk forskning. København, Danmarks pædagogiske Institut, 1975, 195 s.

Garfield, Eugene: Citation Indexing - Its Theory and Application in Science, Technology, and Humanities. New York, John Wiley, 1979.

Garvey, W.D. & Griffith, B.C.: Scientific Communication: Its role in the Conduct of Research and the Creation of Knowledge. American Psychologist. 1971, 26, 349-362.

Garvey, W.D. & Griffith, B.C.: Communication and Information Processing within Scientific Disciplines: Empirical Findings for Psychology. Information Storage and Retrieval, 1972, 8, 123-136.

Gerstberger, P.G. & Allen, T.J.: Criteria Used by Research and Development Engineers in the Selection of an Information Source. Journal of Applied Psychology, 1968, 52, 272-279.

Hjørland, Birger. Psychology and Information Search Strategy: Information Input Overload. Social Science Information Studies, 1984, 4, 143-148.

Hjørland, Birger (1984b). Indlæg om litteratursøgning og databaser på forskningsrådsinitiativernes konference i Middelfart 29 maj 1984. Forventes udgivet af Statens humanistiske Forskningsråd, efterår 1984.

Krørup, Karl og Ivan Boserup: User-oriented indexing. An investigation into the extent to which subject specialists should be used for the indexing of documents by and for professional readers, based on a sample of sociological documents indexed with the help of the PRECIS indexing system. Copenhagen, The Royal Library, 1982. 133 s. (Mimeo).

Large, J.A. The foreign-language barrier: problems in scientific communication. London, Deutsch, 1983. 196 s.

Lau, Birgitte & Karen Risager: Sproglige problemer i informationsvidenskabeligt og biblioteksfagligt arbejde. Rolig Papir 16, 1979, 45-58.

Line, Maurice: The Structure of Social Science Literature as shown by a large-scale citation analysis. Social Science Information Studies, 1981.

Riegelman, Richard K. Studying a study and testing a test: how to read the medical literature. Boston, Little, Brown, 1981. 283 s.

Simkin, James S.: Gestalt Therapy and the Psychological Abstracts. American Psychologist, 1978, 33, 705-706.

Weick, Karl E.: The twigging of overload. (in: People and Information. Ed. by H.P.Pepinsky. New York: Pergamon Press, 1970, p. 67-129).

Westland, Gordon: Current Crisis of Psychology. London, Heinemann, 1978.

Wilson, Patrick: Second hand knowledge. An inquiry into cognitive authority. London, Greenwood Press, 1983. 210 s.

Zachariae, Bobby: Udgrænsningens diskurs: Hinsides psykisk normalitet og patologi. Psykologisk Skriftserie Århus, 1983, vol. 8, no. 9.

Ørnager, Susanne og Mogens Johne: Changes in thesaurus construction caused by the use of boolean searching. (i: 7th International Online Information Meeting London 6-8 December 1983. Oxford: Learned Information, 1983. P. 167-173).

*Skrifter om
Anvendt og
Matematisk
Lingvistik*

11

HUMANISTISK FAKULTET
REPRO AFDELING
Udgivet af Københavns Universitets
Institut for anvendt og matematisk lingvistik
København 1985

INDHOLDSFORTEGNELSE

side i
fortløbende
paginering

Elisabeth Engberg-Pedersen:	
SPROGSTATUS - ordbøger som spejl af et sprogs status (sider B1-B25)	5
Bente Dalgaard, Anne Holmen og Svend Erik Olsen:	
NAVNGIVNING - Nogle studier i personnavne og deres strukturelle og samfundsmæssige kontekst (sider N1-N42)	27
Birger Hjørland:	
FORSKERES SELEKTIONSSTRATEGIER I INFORMATION- SØGNING UNDER "OVERLOAD"-BETINGELSER (sider N1-N42)	69
Charlotte Toubro:	
EDB I FREMMEDSPROGSUNDERVISNINGEN - INTELLIGENT SVAREVALUERING (sider E1-E21)	105
Jesper Hermann:	
SUBJEKTBEGREBET I SPROGPSYKOLOGI (sider L1-L4)...	127
Esther Glahn:	
DATAMATSTØTTET UNDERVISNING I FREMMEDSPROGS- FAGENE (sider D1-D9)	131
Oplysninger om SAML 11	140